IN THE CLAIMS

Please **amend** the claims as indicated:

1-14. (canceled)

15. (currently amended) A method comprising:

reading an HTML document of a web page as an analyzing object;

conducting a temporary block analysis based on a description of HTML tags of the HTML document;

using the HTML tags to temporarily divide the HTML document into blocks;

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements include:

plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type of media used to display the HTML document,

text in the HTML document that is shorter than a maximum predetermined length, and wherein the text appears in the HTML document more than a predetermined frequency,

multiple anchors having a same title,

image tags that only perform a role of punctuation for text in the HTML document, and

multiple text blocks having a same description;

~~deleting~~ defining any block in the HTML document that is deemed to be ~~structurally~~ meaningless as an OBJECT_DELIMITER, wherein a block is deemed to be ~~structurally~~ meaningless if that block [[has]] contains only unnecessary information elements and at least one anchor; and

~~merging relevant information elements in a same block into one composite element~~

crawling only anchors found in blocks that have not been defined as OBJECT_DELIMITERs.

16. (currently amended) The method of claim 15, wherein the ~~unnecessary information~~

~~elements include OBJECT_ANCHORS that have a same title, wherein an OBJECT_ANCHOR describes a correlation between the HTML document and elements in another web page~~ maximum predetermined length is 12 bytes.

17. (currently amended) The method of claim 16, wherein the ~~unnecessary information elements include OBJECT_TEXT_BLOCKS that have a same description of text in a block~~ the predetermined frequency is ten times.

18-20. (canceled)

21. (new) A computer-readable medium encoded with a computer program, wherein the computer program, when executed, performs the steps of:

reading an HTML document of a web page as an analyzing object;

conducting a temporary block analysis based on a description of HTML tags of the HTML document;

using the HTML tags to temporarily divide the HTML document into blocks;

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements include:

plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type of media used to display the HTML document,

text in the HTML document that is shorter than a maximum predetermined length, and wherein the text appears in the HTML document more than a predetermined frequency,

multiple anchors having a same title,

image tags that perform a role of punctuation for text in the HTML document, and

multiple text blocks having a same description;

defining any block in the HTML document that is deemed to be meaningless as an OBJECT_DELIMITER, wherein a block is deemed to be meaningless if that block contains only unnecessary information elements; and

crawling only anchors found in blocks that have not been defined as

OBJECT_DELIMITERs.

22. (new) The computer-readable medium of claim 21, wherein the maximum predetermined length is 12 bytes.

23. (new) The computer-readable medium of claim 21, wherein the predetermined frequency is ten times.

24. (new) A method comprising:

dividing an HTML document into blocks;

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements include:

text in the HTML document that is shorter than a maximum predetermined length, and wherein the text appears in the HTML document more than a predetermined frequency,

multiple anchors having a same title,

image tags that only perform a role of punctuation for text in the HTML document, and

multiple text blocks having a same description;

defining any block in the HTML document that is deemed to be meaningless, wherein a block is deemed to be meaningless if that block contains only the unnecessary information elements and at least one anchor; and

crawling only anchors found in blocks that have not been deemed meaningless for containing only unnecessary information elements.